



Endliche Automaten: Anwendungen: Morphologie

Kursfolien

Karin Haenelt

09.07.2003

Karin Haenelt, FSA-Anwendung:Morphologie

1

Vielen Dank

Für das Aufspüren von Fehlern in früheren Versionen und für
Verbesserungsvorschläge danke ich

Annika Scheffler

Architekturmodelle

- Lexikonerstellung in Teillexika
 - Stämme nach linguistischen Klassen
 - Endungen
 - Morphotaktik und Regeln (bei Zwei-Ebenen-Morphologie)
- Kompilation der Lexika
 - Variante a) in Teilautomaten
 - Variante b) integriertes Gesamtlexikon
- erzeugte Lexikontypen
 - bidirektionale Transduktoren
 - p-sequentielle Transduktoren
jeweils optimiert für Analyse oder Generierung

Zwei-Ebenen-Morphologie mit bidirektionalen Transduktoren

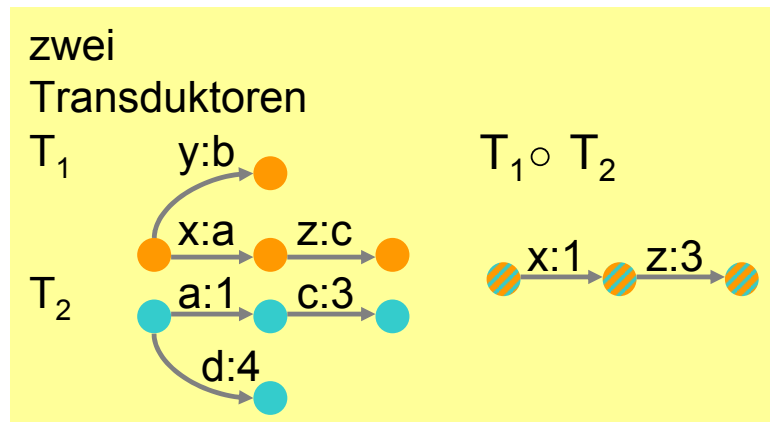
- Lexikonautomat
- Regelautomat, Modellierung orthographischer und phonologischer Regelmäßigkeiten
- Verwendung
 - Variante a:
kaskadierte Transduktoren, d.h. sequentielle Anwendung
 - Variante b:
Komposition von Lexikonautomat und Regelautomat, d.h. gleichzeitige Anwendung

Komposition

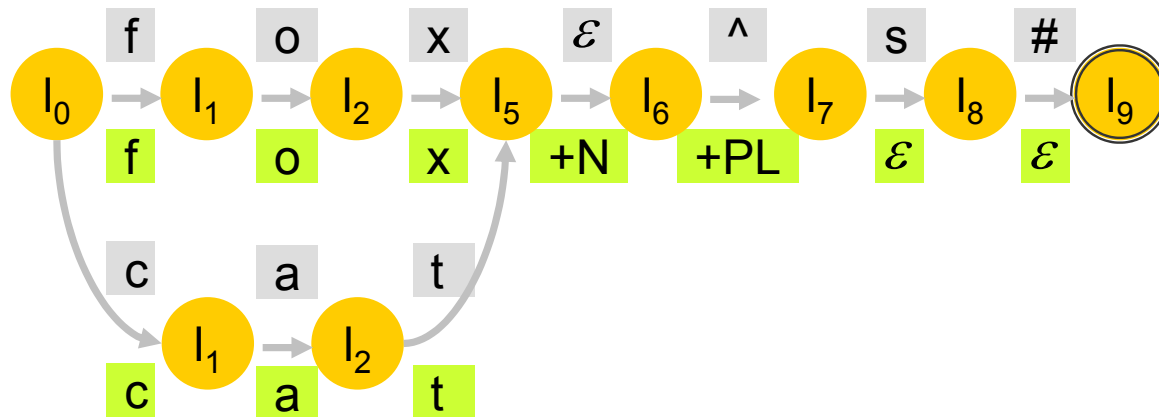
- erfolgt
 - durch Schnittbildung der Ausgabe des ersten Automaten mit der Eingabe des zweiten Automaten, d.h.
- gemeinsame Kanten im Transduktor:

$$\exists c, \text{ so dass } \delta_1(x_a, i : c) = x_b \text{ und}$$

$$\delta_2(y_a, c : o) = y_b$$



Beispiel: Lexikon

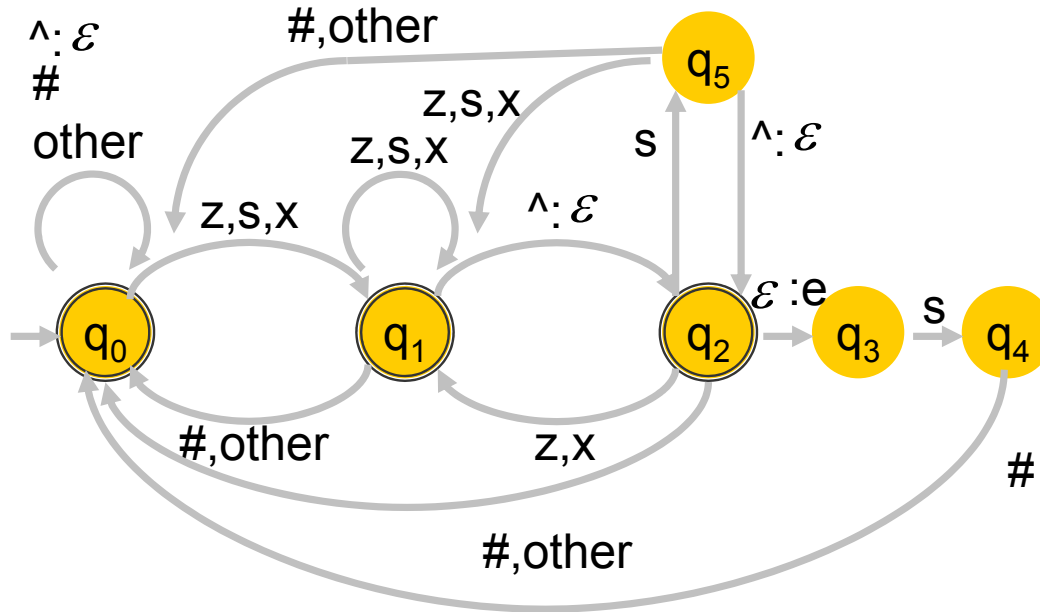


Beispiel: orthographische Regeln

Name	Beschreibung	Beispiel
Konsonatenverdoppelung	Einzelbuchstaben-Konsonant verdoppelt vor <i>-ing/-ed</i>	beg/begging
E-Tilgung	schweigendes e getilgt vor <i>-ing</i> und <i>-ed</i>	make/making
E-Einfügung	e eingefügt nach <i>-s, -z, -x, -ch, -sh</i> vor <i>-s</i>	watch/watches
Y-Ersetzung	<i>-y</i> wird zu <i>-ie</i> vor <i>-s, -i</i> vor <i>-ed</i>	try/tried
K-Einfügung	Verben, die auf Vokal und <i>-c</i> enden, fügen <i>-k</i> ein	panic/panicked

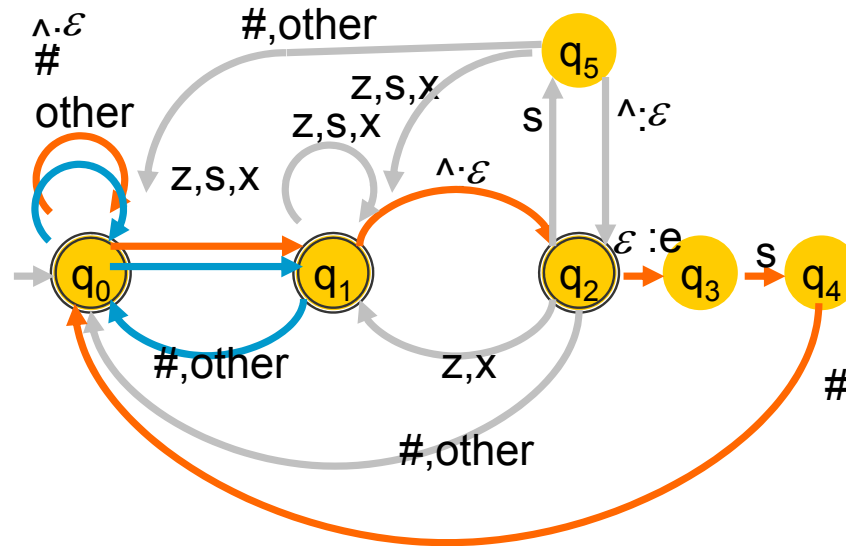
Beispiel: E-Einfügung

$$\varepsilon \rightarrow e / \left\{ \begin{array}{c} x \\ s \\ z \end{array} \right\} \wedge _ s \#$$

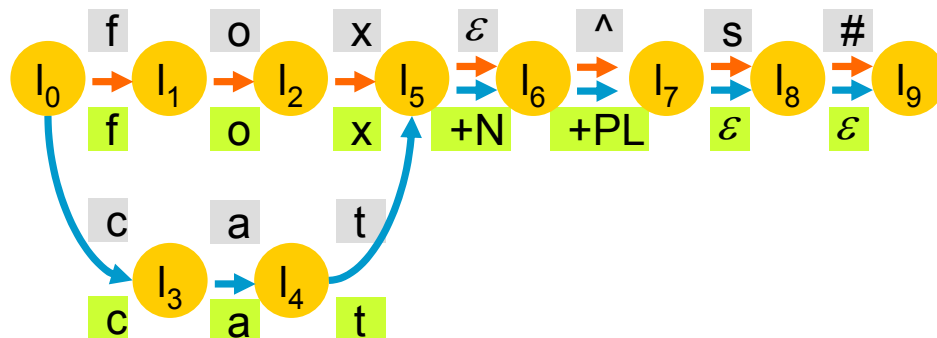


Anwendung der Automaten

R



L



09.07.2003

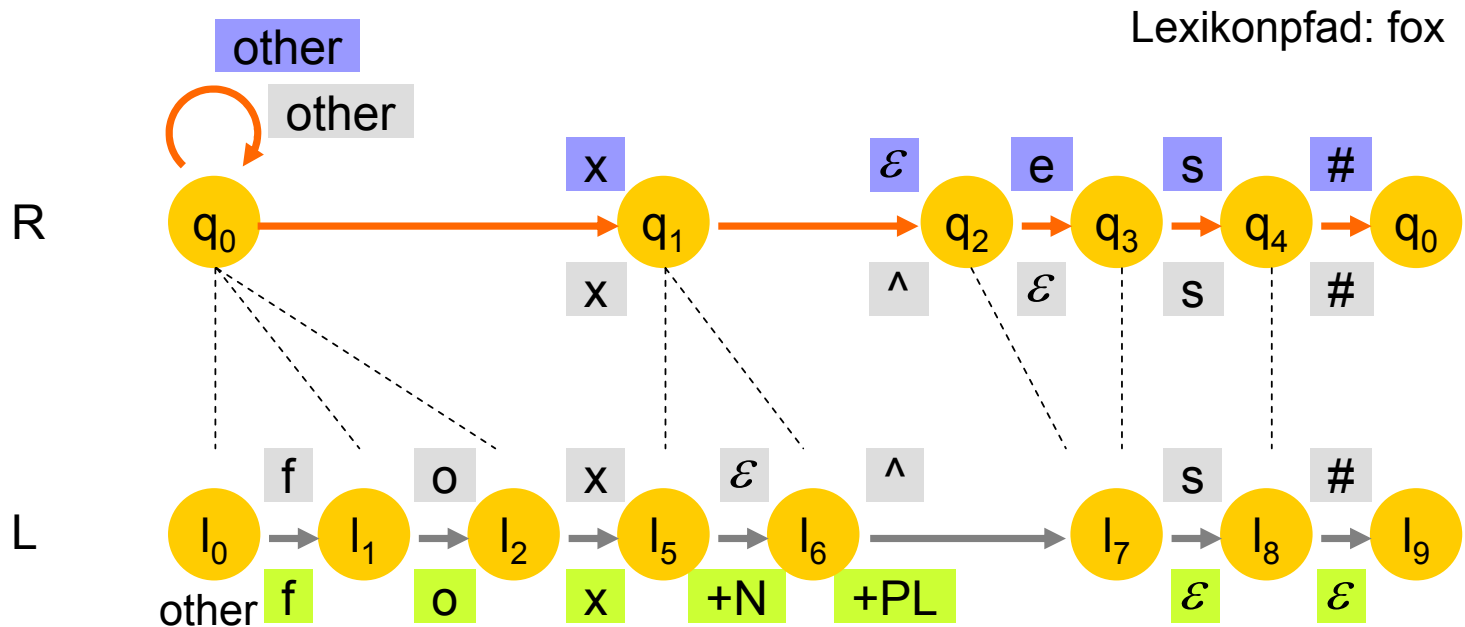
Karin Haenelt, FSA-Anwendung:Morphologie

Komposition

- Komposition
 - von Lexikonautomat und Regelautomat zu einem integrierten Gesamtautomaten
 - erfolgt durch Schnittbildung
- Zustandübergang (q_0 , other, q_0)
am Beginn des Endungsautomaten sorgt dafür,
dass bei Schnittbildung mit dem Lexikonautomaten
der gesamte Lexemteil Element der Durchschnittsmenge wird und
unter Schnittbildung erhalten bleibt
- Algorithmus zur Komposition/Schnittbildung s. Einführung
- Ergebnis der Komposition:
 - Lexikonautomat mit pfadspezifisch komponierten Endungen
 - kann zu Vervielfachungen von Teilen des Regelautomaten führen

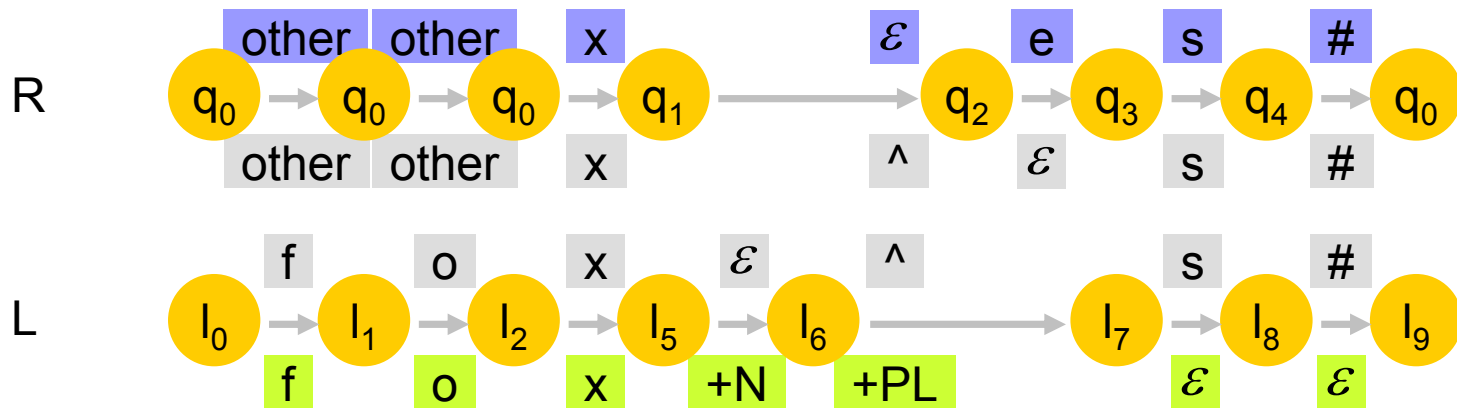
Komposition

Zwischenschritt: Veranschaulichung des Abgleichs von R und L

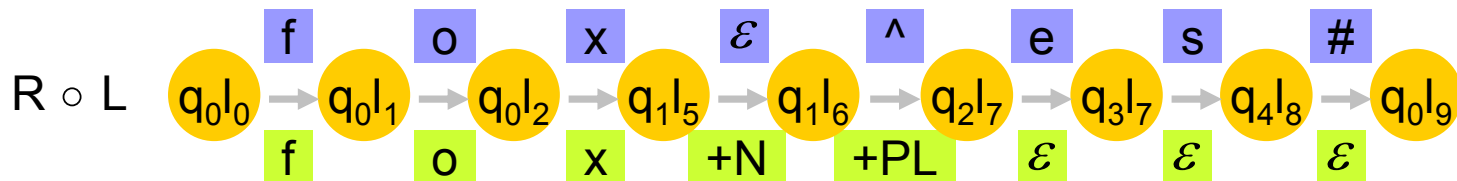


Komposition

Zwischenschritt: Veranschaulichung des Abgleichs von R und L



Ergebnis der Komposition



R	s:s	x:x	z:z	^:ε	ε:e	#	other
q ₀	q ₁	q ₁	q ₁	q ₀	-	q ₀	q ₀
q ₁	q ₁	q ₁	q ₁	q ₂	-	q ₀	q ₀
q ₂	q ₅	q ₁	q ₁	q ₀		q ₀	q ₀
q ₃	q ₄	-	-	-	-	-	-
q ₄	-	-	-	-	-	q ₀	-
q ₅	q ₁	q ₁	q ₁	q ₂	-	-	q ₀

L	a:a	c:c	f:f	o:o	t:t	x:x	+N:ε	+PL:^	ε:s	ε:#
l ₀		l ₃	l ₁							
l ₁				l ₂						
l ₂						l ₅				
l ₃	l ₄									
l ₄					l ₅					
l ₅							l ₆			
l ₆								l ₇		
l ₇									l ₈	
l ₈										l ₉
l ₉										

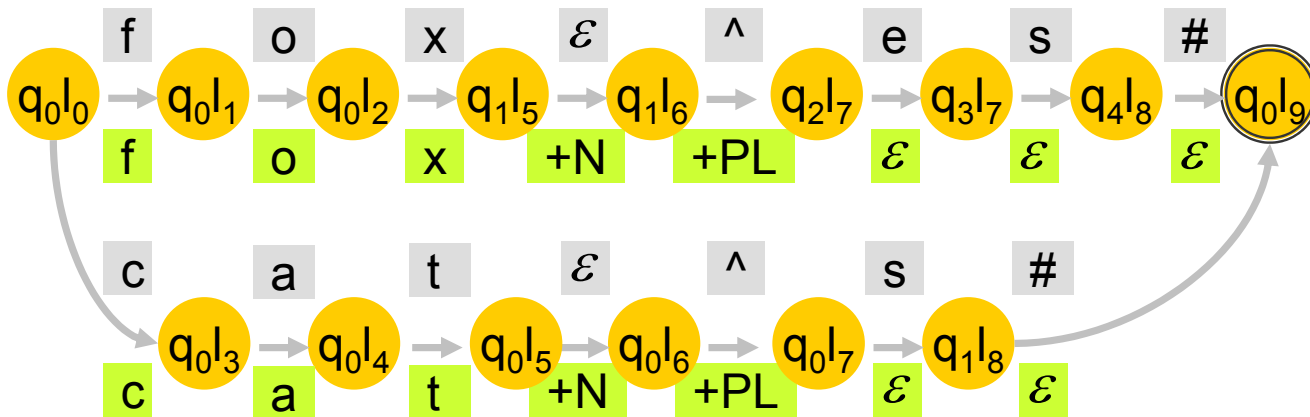
Transitions-

T a b e l l e n

R o L	other	other	other	other	other	x:x	-	^:ε	s:s	ε:e	
	a:a	c:c	f:f	o:o	t:t	x:x	+N:ε	+PL:^	ε:s	-	ε:#
q ₀ l ₀		q ₀ l ₃	q ₀ l ₁								
q ₀ l ₁				q ₀ l ₂							
q ₀ l ₂						q ₁ l ₅					
q ₀ l ₃	q ₀ l ₄										
q ₀ l ₄					q ₀ l ₅						
q ₀ l ₅							q ₀ l ₆				
q ₀ l ₆								q ₀ l ₇			
q ₀ l ₇									q ₁ l ₈		
q ₁ l ₅							q ₁ l ₆				
q ₁ l ₆								q ₂ l ₇			
q ₁ l ₈											q ₀ l ₉
q ₂ l ₇									q ₅ l ₈	q ₃ l ₇	
q ₃ l ₇									q ₄ l ₈		
q ₄ l ₈											q ₀ l ₉
q ₀ l ₉											

09.07.2003

Beispiel: Integrierter Gesamtautomat



Zwei-Ebenen-Morphologie

- Vorteile
 - Erfassung von Regeln reduziert
Kodierungsaufwand
 - erhöht Robustheit gegenüber neuen
Wortbildungen

Integrierter Gesamtautomat

- Vorteile
 - einfache und schnelle Verarbeitung
- Nachteile
 - Größenordnung > 2 GB für Morphologie einer Sprache

FSA-Technik ist nur bei
Hauptspeicherverarbeitung schnell

Literatur

- Beesley Kenneth R. und Lauri Karttunen (2003): *Finite-State Morphology*. Distributed for the Center for the Study of Language and Information. 696 p. (est.). Series: (CSLI-SCL) Studies in Computational Linguistics
- Jurafsky, Daniel; Martin James H. (2000) *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* Upper Saddle River, N.J.: Prentice Hall.